

	<p><b>Calculation of Aquatic Life Use Support Estimates</b> S-001-OWQ-WAP-PRB-20-T-R0 <b>Technical Standard Operating Procedure (TSOP)</b> <b>Office:</b> Office of Water Quality <b>Branch:</b> Watershed Assessment and Planning Branch <b>Section:</b> Probabilistic Monitoring Section</p> <p><b>Last Revised:</b> N/A <b>Revision Cycle:</b> Every 4 years <b>Originally Effective:</b> June 30, 2020</p>
---	--

## **Purpose (of the TSOP)**

This technical standard operating procedure (TSOP) describes the process used to provide statistical estimations of the percentage of perennial stream miles in the Probabilistic Monitoring Program target basin that support aquatic life use following attainment decisions for Indiana's Integrated Water Monitoring and Assessment Report.

## **This TSOP should be used by:**

This TSOP applies to agency staff in the Office of Water Quality (OWQ) Watershed Assessment and Planning Branch (WAPB) Probabilistic Monitoring Section responsible for providing statistical estimations of the percentage of perennial stream miles supporting aquatic life use in the Probabilistic Monitoring Program target basin.

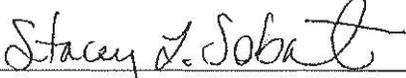
## **Authors**

Paul McMurray, Senior Environmental Manager  
Probabilistic Monitoring Section

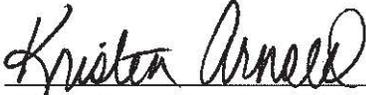
Scott Zello-Dean, Environmental Manager II  
Probabilistic Monitoring Section

## Authorizing Signatures

I approve and authorize this technical standard operating procedure:

  
\_\_\_\_\_  
Stacey Sobat, Section Chief  
OWQ Probabilistic Monitoring Section

07/20/2020  
Date

  
\_\_\_\_\_  
Kristen Arnold, Section Chief  
OWQ Technical and Logistical Services Section

7/23/2020  
Date

  
\_\_\_\_\_  
Marylou Renshaw, Branch Chief  
OWQ Watershed Assessment and Planning Branch

7/23/2020  
Date

This technical standard operating procedure is consistent with agency requirements.

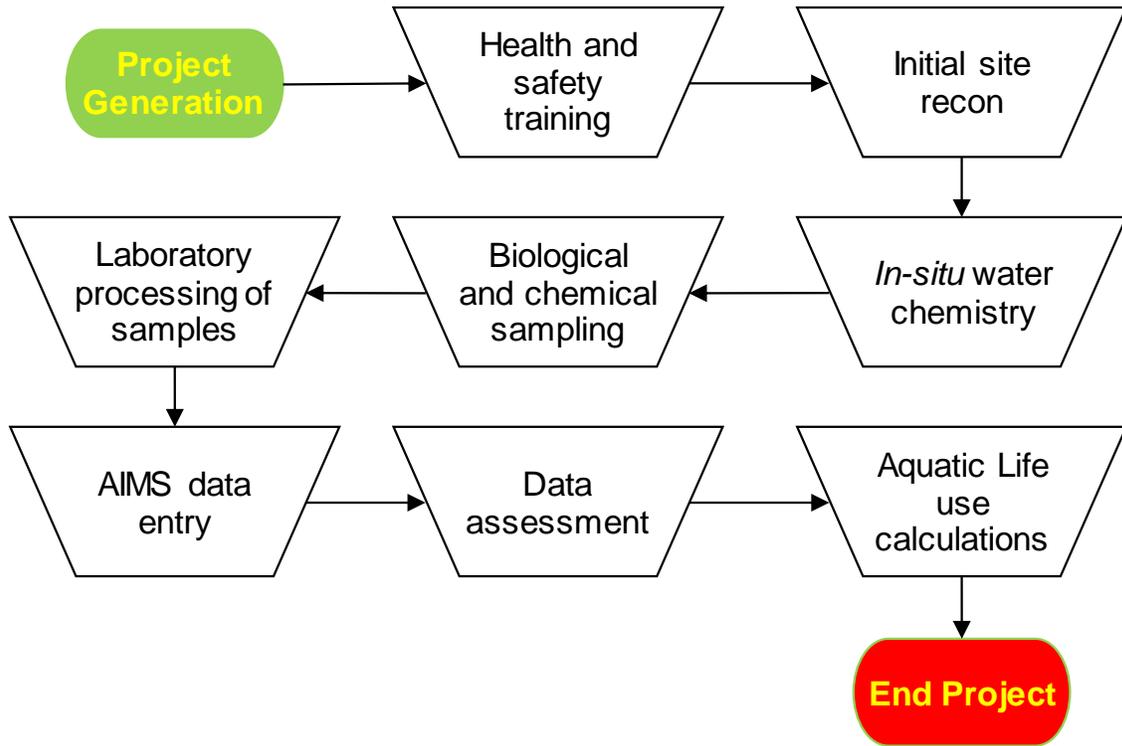
  
\_\_\_\_\_  
Quality Assurance Staff  
IDEM Office of Program Support

24 Jul 2020  
Date

## Table of Contents

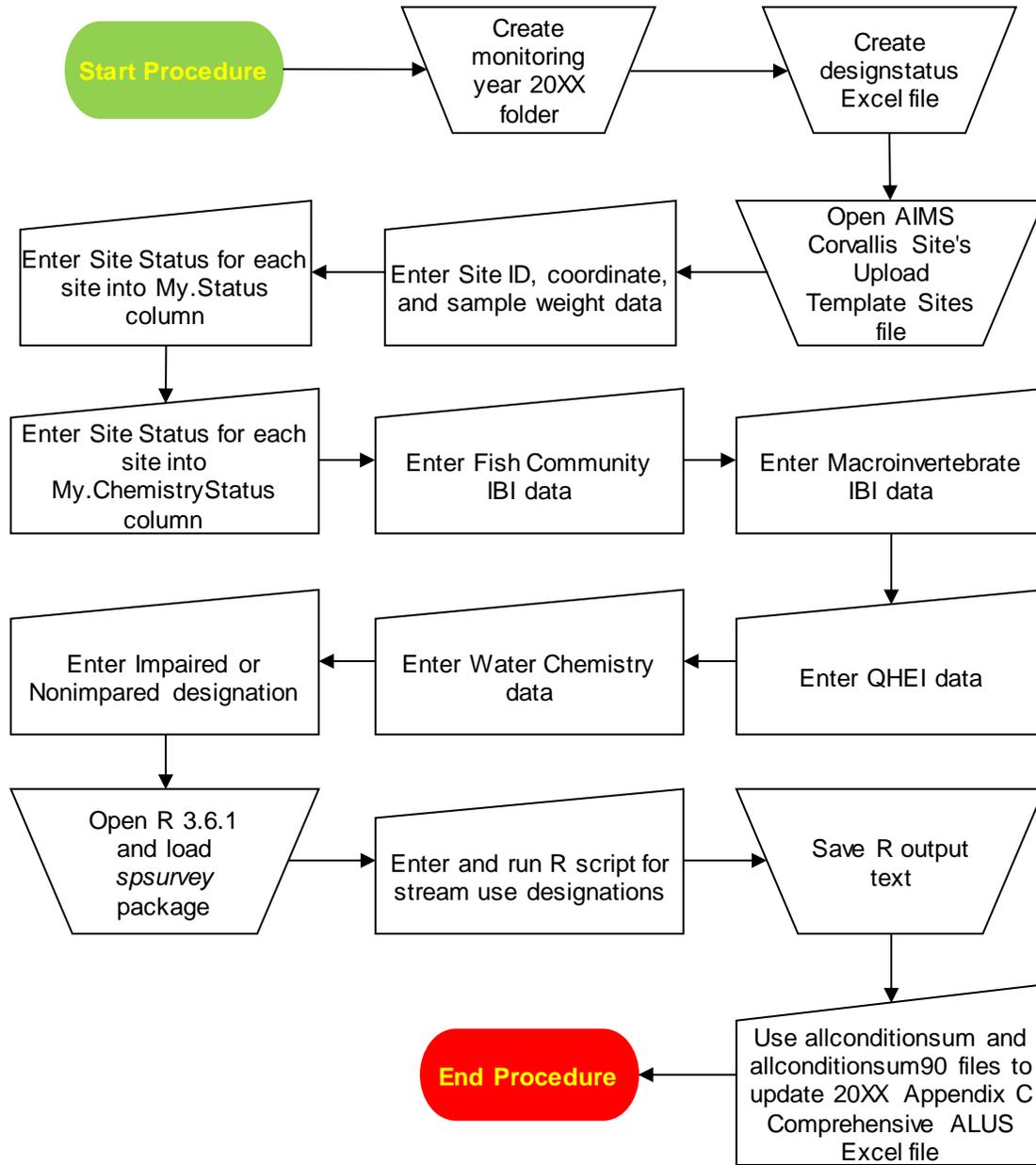
Purpose (of the TSOP) .....	1
This TSOP should be used by: .....	1
Authorizing Signatures .....	2
1.0. Overview Flowchart.....	4
2.0. Procedure .....	5
2.1. Procedural Flowchart:.....	5
2.2. Procedural Steps: .....	5
2.3. Related Technical Issues:.....	16
3.0. Roles.....	18
4.0. List forms, equipment, and/or software to be used .....	19
5.0. Records Management.....	19
6.0. Definitions .....	19
7.0. Quality Assurance / Quality Control.....	20
8.0. References .....	20
9.0. Appendices.....	20

## 1.0. Overview Flowchart



## 2.0. Procedure

### 2.1. Procedural Flowchart:



### 2.2. Procedural Steps:

Step 1. Navigate to S:\IGCN\OWQ\WSP\OWM\Integrated Report Data on the Indiana Department of Environmental Management (IDEM) shared computer drive (S-Drive). Locate and open the folder specific to the probabilistic target basin for which the

percentage of stream miles supporting aquatic life use estimates are being made:

- East Fork White and Whitewater Basins (also known as the Great Miami River basin)
- Lower Wabash and Kankakee Basins
- Tributaries to The Great Lakes and Ohio River Basins
- Upper Wabash Basin
- West Fork White and Patoka Basins

Once the appropriate target basin folder is open, create a new folder entitled Monitoring Year 20XX, where 20XX is the sampling year (e.g., 2018, 2019, 2020, 2021). Save all files created during the calculation process to this new folder.

Step 2. Open a new Excel workbook (.xlsx) and populate row 1 of the first worksheet with the following column headings. Spell each column heading exactly as shown:

Column A – MySite.ID

Column B – Lsite

Column C – SiteName

Column D – Longdd

Column E – Latdd

Column F – My.stratum

Column G – MDCaty

Column H – Weight

Column I – StrahlerOrder

Column J – My.Type

Column K – My.Status

Column L – My.ChemistryStatus

Column M – Fish.Community

Column N – Fish.IBI

Column O – Macro.Community

Column P – Macro.IBI

Column Q – Habitat

Column R – QHEI

Column S – DO

Column T – pH

Column U – Chloride

Column V – Nitrate.and.Nitrite

Column W – Sulfate

Column X – Lead

Column Y – Copper

Column Z – Zinc  
 Column AA – Ammonia  
 Column AB – Cadmium  
 Column AC – Chromium  
 Column AD – Nickel  
 Column AE – Cyanide  
 Column AF – Aluminum  
 Column AG – Selenium  
 Column AH – Nutrients  
 Column AI – Overall.Chemistry  
 Column AJ – Combined.Assessment  
 Name the workbook designstatus.xlsx and save in the Monitoring Year 20XX folder created in Step 1.

- Step 3. Navigate to the S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis20XX folder (where 20XX refers to the sampling year). Open the AIMS\_CorvallisSitesUploadTemplate\_XXXX.xlsx workbook, where XXXX refers to the initials of the probabilistic target basin and the IDEM staff which modified the original workbook. Appendices 1 and 2 contain potential target basin names. When performing calculations for sample years 2011-2016, the workbook will be named AIMS\_DataUploadTemplate\_SitesXXXX.xlsx. Earlier sample years may not be available in this folder and may require locating the original Corvallis site request documentation.
- Step 4. Copy the following data from the DATA worksheet of AIMS\_CorvallisSitesUploadTemplate\_XXXX.xlsx and paste into the designated columns in designstatus.xlsx created in Step 2. Do not copy data from Row 1 (e.g., the column headings) or any rows beyond Row 101.

<b>AIMS_CorvallisSitesUploadTemplate_XXXX</b>	<b>designstatus</b>
EPA Site Identifier	MySite.ID
Site ID	Lsite
Stream Name	SiteName
Longitude (decimal degree)	Longdd
Latitude (decimal degree)	Latdd
Weight Category	MDCaty
DesignWeightSurvey	Weight
Strahler Order	StrahlerOrder

Complete the following modifications to designstatus.xlsx:

- In the MDCat column, rename entries of 1 to 1st, 2 to 2nd, 3 to 3<sup>rd</sup>, and 4 to 4th+
- The My.stratum column contains a number representing the probabilistic target basin chronological sampling order. Based on the year in which sampling was conducted, enter the appropriate number from the My.stratum column in Appendices 1 or 2 into all of the rows of My.stratum column.
- The My.Type column refers to the name of the probabilistic target basin. Enter the name of the probabilistic target basin from the My.Type column in Appendices 1 or 2 into all of the rows of the My.Type column.
- If the Site ID column of AIMS\_CorvallisSitesUploadTemplate\_XXXX.xlsx is blank, fill in the Lsite column using the information sources described in Step 5.

Step 5. Locate the work plan created for the probabilistic target basin in the S:\IGCN\OWQ\WSP\OWM\Document\Workplans\20XX folder, where 20XX refers to the sampling year. Within the work plan, locate the table listing all potential sites for the probabilistic target basin. In this table, the numbers in the Site # column correspond to the last three digits of the designstatus.xlsx's MySite.ID column.

Copy the information from the AIMS Site Name column in the probabilistic work plan potential site table into the designstatus.xlsx's LSite column. The work plan table also contains each site's Site Status (i.e., Target, Approved, or Landowner Denial) determined during reconnaissance activities. During the sampling process, sites may be subsequently rejected if, for example, the site goes dry.

Using the spreadsheets and notes compiled for and taken during the Probabilistic Monitoring Program aquatic life use assessments, enter the appropriate code for each site into the designstatus.xlsx's My.status column. Spreadsheets and notes are available from the Integrated Report coordinator; staff responsible for the fish, macroinvertebrate, and water chemistry sampling programs; Probabilistic Monitoring Program site sampling file folders; and AIMSII database queries. Site status designations derived post sampling from the Probabilistic

Monitoring Program aquatic life use assessment meetings' notes, Probabilistic Monitoring Program site sampling file folders, and AIMSII database queries supersede the Probabilistic Monitoring Program work plan's initial designations.

TS – target site; a site where both fish and macroinvertebrate communities and three rounds of water chemistry data were collected.

SBNC – sampled biology, no chemistry; a site where fish and macroinvertebrate communities were collected, but not three rounds of water chemistry sampling.

SFNCM – sampled fish, no chemistry or macroinvertebrates; a site where only fish community sampling occurred, but not macroinvertebrate community and three rounds of water chemistry sampling.

SMNCF – sampled macroinvertebrates, no chemistry or fish; a site where only macroinvertebrate community sampling occurred, but not fish community and three rounds of water chemistry sampling.

SCNB – sampled chemistry, no biology; a site where three rounds of water chemistry data sampling occurred, but not fish and macroinvertebrate communities sampling.

SCFNM – sampled chemistry and fish, no macroinvertebrates; a site where three rounds of water chemistry data and fish community sampling occurred, but not macroinvertebrate community sampling.

SCMNF – sampled chemistry and macroinvertebrates, no fish; a site where three rounds of water chemistry data and macroinvertebrate community sampling occurred, but not fish community sampling.

PB – physical barrier; access to site is blocked (e.g., logjam, beaver dam).

OT – Other; a site rejected for safety reasons.

NN – not needed; an approved site not sampled or potential target site which could not be assessed by the reconnaissance deadline.

LD – landowner denial; a site in which the landowner does not grant IDEM permission to collect samples either during the

reconnaissance process or at any point during the sampling season.

NT – nontarget; a site rejected for a reason other than landowner denial. Potential rejection reasons include the site is located in a backwater, oxbow, wetland, a tiled section of the stream, or the stream was dry during reconnaissance or subsequently went dry during the sampling season.

Step 6. Copy site status designations from the designstatus.xlsx's My.status column to the designstatus.xlsx's My.ChemistryStatus column. Change site status designations as follows:

<b>My.Status designation</b>	<b>My.Chemistry.Status designation</b>
<b>TS, LD, NT, PB, OT and NN</b>	<b>TS, LD, NT, PB, OT and NN (No change)</b>
<b>SCNB, SCFNM and SCMNF</b>	<b>TS</b>
<b>SBNC, SFNCM and SMNCF</b>	<b>NT</b>

Step 7. Use the data described in Step 5 to enter the Index of Biotic Integrity (IBI) score in the designstatus.xlsx's Fish.IBI column, for fish community samples collected at probabilistic sample sites. If more than one IBI score is available for a site (e.g., a revisit sample collected for QA/QC purposes), use the IBI score used to derive the pass or fail status of the site. If probabilistic monitoring staff used BPJ (best professional judgement) to pass an otherwise failing sample use the status indicated in notes taken during the Probabilistic Monitoring Program aquatic life use assessment meetings.

Use the Fish.IBI column scores to enter one of the following in the designstatus.xlsx's Fish.Community column:

- I (impaired) for an IBI score less than 36.
- NI (not impaired) for an IBI score equal to or greater than 36.
- NA (not applicable) if a fish community IBI score is not available for the site.

Step 8. Use the data described in Step 5 to enter the macroinvertebrate Index of Biotic Integrity (mIBI) score in the designstatus.xlsx's Macro.IBI column for macroinvertebrate community samples collected from probabilistic sample sites. If more than one mIBI score is available (e.g. a duplicate sample collected for QA/QC purposes), use the mIBI score used to derive the site's pass or fail status. If probabilistic monitoring staff used BPJ (best professional judgement) to pass an otherwise failing sample use

the status indicated in notes taken during the Probabilistic Monitoring Program aquatic life use assessment meetings.

Use the designstatus.xlsx's Macro.IBI column scores to enter one of the following in the designstatus.xlsx's Macro.Community column:

- I (impaired) for a mIBI score less than 36.
- NI (not impaired) for a mIBI score equal to or greater than 36.
- NA (not applicable) if a macroinvertebrate community mIBI score is not available for the site.

Step 9. Use the data described in Step 5 to enter the Qualitative Habitat Evaluation Index (QHEI) score associated with the fish community sample entered in Step 5 in the designstatus.xlsx's QHEI column. When using a fish community revisit sample to derive a site's pass or fail status, use the QHEI score associated with the revisit sample.

Use the designstatus.xlsx's QHEI column scores to enter one of the following in the designstatus.xlsx's Habitat column:

- I (impaired) for a QHEI score less than 51.
- NI (not impaired) for a QHEI score equal to or greater than 51.
- NA (not applicable) if a QHEI score is not available for the site.

Step 10. Use the data described in Step 5 to enter the following impairment statuses for chemical parameters into the designstatus.xlsx's DO (dissolved oxygen), pH, Chloride, Nitrate.And.Nitrite, Sulfate, Lead, Copper, Zinc, Ammonia, Cadmium, Chromium, Nickel, Cyanide, Aluminum, Selenium, and Nutrients columns:

- I (impaired)
- NI (not impaired)
- NA (not applicable) if three rounds of water chemistry sampling were not conducted at a site, which is indicated in notes taken during the Probabilistic Monitoring Program aquatic life use assessment meetings.

If all of a site's chemical parameter results are NI, enter NI in the designstatus.xlsx's Overall.Chemistry column. If any one of the individual chemical parameter results for a site are I, enter I into the Overall.Chemistry column. If three rounds of water chemistry

sampling were not conducted, enter NA in the Overall.Chemistry column.

- Step 11. If all of the results in the designstatus.xlsx's Fish.Community, Macro.Community, and Overall.Chemistry columns for a site are NI, enter NI in the designstatus.xlsx's Combined.Assessment column. If any one of the columns' results is I, enter I in the Combined.Assessment column. If any one of the columns' results is NA, enter NA in the Combined.Assessment column.

Save the designstatus.xlsx as both an Excel Worksheet (.xlsx) and as a Text (Tab delimited) (.txt) file within the Monitoring Year 20XX folder created in Step 1. Exit Excel.

- Step 12. Open R 3.6.1 and load the *spsurvey* package. Appendix 3 contains instructions on downloading and installing R 3.6.1 and installing and loading the *spsurvey* package.
- Step 13. Enter the following italicized commands into the R Console at the command line prompt. The 2014 Probabilistic Monitoring Program project in the Great Miami River basin (Appendix 4) is used as an example, for the remainder of this TSOP. These command lines may be typed in, exactly as written below, or copied and pasted into the command line of the R Console. Enter commands singly and press Enter after each command line is complete. If a command is written in this TSOP over multiple lines, copy and paste each line of text individually. Additional instructions or explanations for a command follow the # sign and do not alter the command if entered into the R Console. Sections of the commands in bold font are dependent upon the target basin's My.type name and framesize (Appendices 1 and 2).

```
library(spsurvey) # Loads the spsurvey package.
```

```
framesize <- 2330.09 # The framesize value varies dependent upon the basin and the year sampled. U.S. EPA documentation letters provide a basin's framesize value, summarized in Appendices 1 and 2.
```

```
names(framesize) <- 'GreatMiamiRiver' # The names value varies dependent upon the basin sampled. U.S. EPA documentation letters provide a basin's names value, summarized in Appendices 1 and 2.
```

```
setwd("S:/IGCN/OWQ/WSP/OWM/Integrated Report Data/East Fork White and Whitewater Basins/Monitoring Year 2014") #
```

Change East Fork White and Whitewater Basins and Monitoring Year 2014 to the correct basin name and sampling year. Ensure that the words in the basin name are separated by spaces and spelled correctly. U.S. EPA documentation letters provide a basin's name and sampling year, summarized in Appendices 1 and 2.

```
DesignStatus <- read.table("designstatus.txt",header = TRUE,
sep = "\t", na.strings = "NA", dec = ".", strip.white = TRUE)
names(DesignStatus)

tst <- !DesignStatus$My.Status=='NN'
DesignStatus <- data.frame(DesignStatus[tst,])
tmp <- marinus(DesignStatus$Latdd,DesignStatus$Longdd)
DesignStatus$xmarinus <- tmp[, 'x']
DesignStatus$ymarinus <- tmp[, 'y']
sites <- rep(TRUE,nrow(DesignStatus))

DesignStatus$final.wt <-
adjwgt(sites,DesignStatus$Weight,DesignStatus$My.Type,frame
size) # These three lines should be entered onto one R
command line with "framesize" as a whole word.

sum(DesignStatus$Weight)

sum(DesignStatus$final.wt) # The command's output should
equal the framesize entered into the first command line.

write.table(DesignStatus,
file="designstatus.csv",sep="," ,row.names=FALSE) # Creates a
Comma Separated Values (.csv) Excel file titled designstatus.csv
in the Monitoring Year 20XX folder.

designstatus <-
read.table('designstatus.csv',sep=',',header=TRUE)

sites <-
data.frame(siteID=designstatus$MySite.ID,Use=rep(TRUE,nrow(
designstatus)))

subpop <- data.frame( siteID=designstatus$MySite.ID,
GreatMiamiRiver=designstatus$My.Type) # Change
GreatMiamiRiver to the correct basin name. U.S. EPA
documentation letters provide the basin name, summarized in
Appendices 1 and 2.
```

```
dsgn <- data.frame(  
siteID=designstatus$MySite.ID, stratum=designstatus$My.stratum,  
wgt=designstatus$final.wt, xcoord=designstatus$xmarinus, ycoord=designstatus$ymarinus) # These three lines should be  
entered onto one R command line with "stratum" and "ycoord" as  
whole words.
```

```
designstatus$StatusTNT <- designstatus$My.Status
```

```
levels(designstatus$StatusTNT) <-  
list(T=c('TS', 'SBNC', 'PB', 'LD', 'SCNB', 'SFNCM', 'SCMNF', 'SCFNM'),  
NT=c('NT')) # If a new sample type abbreviation is created (see  
Step 10), enter it into this command line bracketed by  
apostrophe marks.
```

```
data <-
```

```
data.frame(siteID=designstatus$MySite.ID, CombinedStatus=designstatus$My.Status, StatusTNT=designstatus$StatusTNT) #  
These three lines should be entered onto one R command line  
with "designstatus" as a whole word.
```

```
framelen <- c('GreatMiamiRiver'=2330.09) # Change  
GreatMiamiRiver and 2330.09 to the correct basin name and  
framesize. U.S. EPA documentation letters provide the basin's  
name and framesize, summarized in Appendices 1 and 2.
```

```
sitestatussum <- cat.analysis(sites, subpop, dsgn, data,  
popsize=list(GreatMiamiRiver=framelen)) # Change  
GreatMiamiRiver to the correct basin name. U.S. EPA  
documentation letters provide the basin's name, summarized in  
Appendices 1 and 2.
```

```
write.table(sitestatussum, file="sitestatussum.csv",  
sep="," , col.names=NA) # Creates an Excel file titled  
sitestatussum.csv in the Monitoring Year 20XX folder  
summarizing the number of sites found in each of the category  
types listed in Step 10.
```

```
sites <-
```

```
data.frame(siteID=designstatus$MySite.ID, Use=designstatus$My.Status=='TS') # These three lines should be entered onto one  
R command line with "My.Status" as a whole word.
```

```
data <-
```

```
data.frame(siteID=designstatus$MySite.ID, Status=designstatus$Combined.Assessment)
```

```
conditionsum <- cat.analysis(sites,subpop,dsgn,data)
write.table(conditionsum, file="allconditionsum.csv", sep =
",",col.names=NA) # Creates an Excel file titled
allconditionsum.csv in the Monitoring Year 20XX folder.
conditionsum90 <- cat.analysis(sites,subpop,dsgn,data,conf=90)
write.table(conditionsum90, file="allconditionsum90.csv", sep =
",",col.names=NA) # Creates an Excel file titled
allconditionsum90.csv in the Monitoring Year 20XX folder.
savehistory("Great Miami 2014 - R Analyses.txt") # Creates a
text file of the commands entered into the current R session
which can be opened using Notepad, WordPad, or Word.
Change Great Miami 2014 to the correct basin name and
sampling year. U.S. EPA documentation letters provide the
basin's name, summarized in Appendices 1 and 2.
savehistory("Great Miami 2014 - R Analyses.Rdata") # Creates
an R Workspace data file. To open the file through the R
Console, click File, then Open Script. Locate the file directory,
select the file (may require searching for All files(*.*)). Click
Open. Change Great Miami 2014 to the correct basin name and
sampling year. U.S. EPA documentation letters provide the
basin's name, summarized in Appendices 1 and 2.
```

- Step 14. Open a new Word document. Highlight all text within the R Console (commands and output). Right click and select Copy. Within the Word document, right click and select Paste. Save the Word document as Basin Year – R Analyses, where the Basin and Year reflect the sampled probabilistic target basin and the sampling year. Exit the Word document and the R Console.
- Step 15. Navigate to S:\IGCN\OWQ\WSP\OWM\Integrated Report Data. Locate and open the folder 20XX Integrated Report as Submitted to U.S. EPA where 20XX refers to the year in which the report is submitted. Appendix 3 contains the appropriate report for the year in which the basin of interest was sampled. Open 20XX Appendix C Comprehensive ALUS.xls where 20XX refers to the year in which the report is submitted to U.S. EPA. If the necessary 20XX Integrated Report... folder is missing, create a new folder in S:\IGCN\OWQ\WSP\OWM\Integrated Report Data with the title 20XX Integrated Report as Submitted to U.S. EPA. 20XX refers to the year in which the report is submitted to U.S. EPA. Copy the 20XX Appendix C Comprehensive ALUS.xls from

the previous 20XX Integrated Report... folder into this location and change the 20XX in the file name to the year in which the report is submitted.

Step 16. Open allconditionsum.csv and allconditionsum90.csv created in Step 13 and saved in the Monitoring Year 20XX folder created in Step 1. Cells G2 and G3 in the Estimate.P column of both files represent the probabilistic target basin's percentage of stream miles not attaining and attaining aquatic life uses, respectively. In the 20XX Appendix C Comprehensive ALUS.xls opened in Step 15, enter these values, rounded to the nearest whole number, into the sampled target basin's row in columns G (% Attaining) and H (% Not Attaining).

Calculate the 95% confidence interval by subtracting cell I2 from cell G2 in allconditionssum.csv. Enter the result rounded to the nearest whole number into the sampled target basin's row in column J (Confidence Interval (%)) of the 20XX Appendix C Comprehensive ALUS.xls.

Calculate the 90% confidence interval by subtracting cell I2 from cell G2 in the allconditionssum90.csv. Enter the result rounded to the nearest whole number into the sampled target basin's row in column J (Confidence Interval (%)) of the 20XX Appendix C Comprehensive ALUS.xls.

Update Column I (Confidence Level (%)) of the 20XX Appendix C Comprehensive ALUS.xls file for the sampled target basin's row to reflect the number (n=XX) of probabilistic sites considered Target sites. The number is found in cell F4 of the allconditionsum.csv and allconditionsum90.csv.

Update Column E (Year Assessed) of the 20XX Appendix C Comprehensive ALUS.xls for the sampled target basin's row to reflect the year in which the basin's assessment meetings occurred, usually the year after sampling. Save 20XX Appendix C Comprehensive ALUS.xls and close all open Excel files.

### **2.3. Related Technical Issues:**

#### **A. Health and Safety Warnings:**

Not applicable. This TSOP does not involve travel or field work. All steps are to be completed in an office setting on a computer workstation.

#### **B. Cautions:**

- The R script (detailed in Step 13) used to provide estimations of the percentage of perennial stream miles in the Probabilistic Monitoring Program target basin supporting or not supporting aquatic life use was provided to IDEM by the U.S. EPA. Certain sections of the R script change, with the location and year in which sampling was conducted or the categories used.
- Review notes made during Probabilistic Monitoring Program assessments to ensure the impairment designation made during assessments matches the impairment entry made in the above fish or macroinvertebrate columns. Impairment designations using best professional judgment (BPJ) may pass a site with a failing IBI or mIBI score. BPJ designated sites may occur when staff believe the presence of one or more rare, sensitive, or intolerant taxa may provide justification for passing the sample, in spite of a failing IBI or mIBI score.
- The framesize of the target basin varies slightly between sampling years due to the increased resolution in the National Hydrography Dataset (NHD) used to calculate the values. Always use the framesize associated with the U.S. EPA site draw for the sampling year and target basin. The location of U.S. EPA documentation letters with probabilistic site draws are summarized in Appendix 2.
- Ensure the working directory (i.e. the Monitoring Year 20XX folder specified in the third command line of the R script) is correct for the area for which calculations are being made. Selecting the wrong directory can result in files produced during previous aquatic life use calculations (i.e. designstatus.xlsx) being accidentally overwritten.
- Ensure the working directory path command line uses backward slashes (“/”) or two forward slashes(“\\”) to indicate folder divisions, instead of the default single forward slash (“”).

C. Interferences:

Not applicable.

D. Calibration:

- Staff determining aquatic life use estimations for the first time should attempt to duplicate the results of previous estimations obtained by trained staff.

E. Troubleshooting:

- R is a free software environment for statistical computing and graphics available at <https://www.r-project.org/>. The R website provides manuals (<https://cran.r-project.org/manuals.html>) and FAQs (<https://cran.r-project.org/doc/FAQ/R-FAQ.html>) related to installing and operating R.
- Additional information about the *spsurvey* package can be found at <https://cran.r-project.org/web/packages/spsurvey/index.html>
- If a command is entered into R incorrectly, an error message will appear in the line below the command. Pressing the up (↑) arrow on the keyboard scrolls through the previous command lines. Fix and resubmit the command by pressing Enter. Common sources of errors include the addition of spaces between portions of the command, incorrect capitalization of a letter or misentered alphanumeric characters. In general, a space should always be inserted before and after an assignment operator (<-).
- If R Analyses.txt and R Analyses.Rdata contain lines of code from a previous R session, search the previous working directory. Find the location by entering *getwd()* into the R command line. Delete the file titled .Rhistory in the working directory. Then run the Step 13 commands again.

### 3.0. Roles

#### 3.1. Responsibilities:

- A. Probabilistic Monitoring Program Coordinator
  1. Provide a calculation of target basin aquatic life use estimations.
  2. Create and maintain notes from the Aquatic Life use assessment meetings, and make the notes available for use in calculating target basin aquatic life use estimations.
- B. IDEM Integrated Report Coordinator
  1. Create and maintain notes from the aquatic life use assessment meetings and make the notes available for use in calculating target basin aquatic life use estimations.
- C. Probabilistic Monitoring Section Chief
  1. Provide a second calculation of target basin aquatic life use estimations.
  2. Designate a Probabilistic Monitoring Section staff to conduct a second calculation of target basin aquatic life use estimations.

#### 3.2. Training requirements:

- A. AIMSII database management

1. Probabilistic Monitoring Program Coordinator
  2. Probabilistic Monitoring Section Chief
- B. R statistical program use
1. Probabilistic Monitoring Program Coordinator
  2. Probabilistic Monitoring Section Chief

#### **4.0. List forms, equipment, and/or software to be used**

##### **4.1. Forms**

Not applicable.

##### **4.2. Equipment:**

- A. Computer workstation

##### **4.3. Software:**

- A. AIMSII database
- B. R statistical program

#### **5.0. Records Management**

The process outlined in the TSOP results in the creation of several Excel files (designstatus, sitestatussum, allconditionsum, allconditionsum90, 20XX Appendix C Comprehensive ALUS), text files and Word documents. Electronic files are stored in S:\IGCN\OWQ\WSP\OWM\Integrated Report Data folders specific to the target basin and sampling year. Data generated by this process is included in the Integrated Report submitted to U.S. EPA. Physical copies of the Aquatic Life Use Assessment meeting notes are kept with the Probabilistic Target Basin sampling site folders. These folders are maintained at the IDEM WAPB offices until they are scanned and uploaded into Virtual File Cabinet.

#### **6.0. Definitions**

- 6.1.** “Assessment Information Management System database (AIMSII database)” – IDEM database containing information related to water chemistry; aquatic habitat; macroinvertebrate, fish, and algae communities; fish tissue analyses; sediments; and E. coli bacteria data collected by agency staff from watershed sampling events.
- 6.2.** “L-Site” – A unique Site ID generated by the AIMSII database for each sampling site, and linked with location information (stream name, description, latitude, longitude, county, HUC12, HUC14, etc.).
- 6.3.** “Quality control” – The overall system of technical activities that measures the attributes and performance of a process, item, or service against

defined standards to verify that they meet the stated requirements established by the customer; operational techniques and activities that are used to fulfill requirements for quality. In other words, QC involves measuring the "thing produced" against a standard to ensure it is a quality product that meets the identified need.

- 6.4. "Reconnaissance" – Obtaining information about a site through visual observations, investigating routes to safely access the site, and gathering property owner information and access permission.
- 6.5. "Site number" – The designated program area specific numeral given to an environmental location of interest.

## 7.0. Quality Assurance / Quality Control

New WAPB staff should review this TSOP prior to calculating statistical estimations of the percentage of perennial stream miles in the Probabilistic Monitoring Program target basin supporting or not supporting aquatic life use. After the estimations have been completed by the Probabilistic Monitoring Program coordinator, a second WAPB staff member (such as the Probabilistic Monitoring Section chief or designee) reviews the data within [designstatus.xlsx](#) and calculates the estimate of stream miles supporting aquatic life use. Discuss and resolve any discrepancies before the data is finalized.

## 8.0. References

- 8.1. (IDEM 2020). Office of Water Quality Notice of Public Comment Period for the 2020 List of Impaired Waters and Consolidated Assessment and Listing Methodology under Section 303(d) of the Clean Water Act Appendix 1: IDEM's 2020 Consolidated Assessment and Listing Methodology (CALM). Office of Water Quality, Indiana Department of Environmental Management, Indianapolis, Indiana.
- 8.2. (R Core Team 2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

## 9.0. Appendices

Appendix 1 – Summary of previous aquatic life use assessments from monitoring years 1996 – 2018

Appendix 2 – Summary of locations of Probabilistic Monitoring Program documentation files

Appendix 3 – Installing R 3.6.1 and loading the *spsurvey* package

Appendix 4 – R output from the calculation of aquatic life use support estimates  
for the 2014 Great Miami River Basin Probabilistic Monitoring project

## Appendix 1 – Summary of previous aquatic life use assessments from monitoring years 1996 – 2018

Basin Assessed	HUCs Sampled	Basin Size (miles)	Round	My.stratum	My.Type	Frame Size (km)	Year Sampled	Year Assessed	Data Types Used in Assessment	Percent Attaining	Percent Not Attaining	N sampled	95% Confidence Interval
West Fork White River	05120201 05120202 05120203	4275	1st	1	West Fork White River	5974.82	1996	1997	Biological	47%	53%	31	±17%
		4275	2nd	7	WestForkWhiteRiver	5974.82	2001	2003	Biological Chemical	67%	33%	36	±15%
		4275	3rd	1	WestForkWhiteRiver	6074.201	2006	2008	Biological Chemical	51%	49%	38	±17%
		3774	4th	27	WestForkWhiteRiver	6074.201	2011	2013	Biological Chemical	71%	29%	38	±15%
East Fork White River	05120204 05120205 05120206 05120207 05120208	4856	1st	2	East Fork White River	6802.39	1997	1998	Biological	40%	60%	32	±17%
		4856	2nd	9	EastForkWhite	6802.385	2002	2004	Biological Chemical	73%	27%	34	±11%
		4856	3rd	3	EastForkWhiteRiver	6959.485	2007	2009	Biological Chemical	35%	65%	32	±15%
		4325	4th	28	EastForkWhite	6960.41	2013	2014	Biological Chemical	61%	39%	37	±15%
Great Miami River	05080001 05080002 05080003	1621	1st	2	Whitewater	2270.02	1997	1998	Biological	77%	23%	19	±20%
		1621	2nd	8	Whitewater	2270.018	2002	2004	Biological Chemical	90%	10%	26	±10%
		1621	3rd	4	Whitewater	2330.508	2007	2009	Biological Chemical	70%	30%	38	±14%
		1621	4th	30	GreatMiamiRiver	2330.09	2014	2015	Biological Chemical	78%	22%	38	±13%
Upper Wabash River	05120101 05120102 05120103 05120104 05120105 05120106 05120107	6632	1st	3	Upper Wabash	9548.06	1998	1999	Biological	52%	48%	48	±16%
		6632	2nd	10	UpperWabash	9548.056	2003	2005	Biological Chemical	25%	75%	37	±14%
		6632	3rd	22	UpperWabash	9904.22	2008	2010	Biological Chemical	29%	71%	36	±13%
		6632	4th	31	UpperWabashRiver	9904.92	2015	2016	Biological Chemical	56%	44%	36	±16%
Lower Wabash River	05120108 05120109 05120110 05120111 05120113	5306	1st	4	Lower Wabash	7601.42	1999	2002	Biological	64%	36%	39	±15%
		5306	2nd	12	LowerWabash	7601.418	2004	2006	Biological Chemical	58%	42%	34	±15%
		5306	3rd	24	LowerWabash	7474.01	2009	2011	Biological Chemical	59%	41%	38	±17%
		5306	4th	32	LowerWabashRiver	7625.33	2016	2017	Biological Chemical	43%	57%	37	±14%

**Appendix 1 – Summary of previous aquatic life use assessments from monitoring years 1996 – 2018 (cont.)**

Basin Assessed	HUCs Sampled	Basin Size (miles)	Round	My.stratum	My.Type	Frame Size (km)	Year Sampled	Year Assessed	Data Types Used in Assessment	Percent Attaining	Percent Not Attaining	N sampled	95% Confidence Interval
Kankakee River	07120001 07120002 07120003	3958	1st	4	Kankakee	5606.33	1999	2002	Biological	47%	53%	41	±16%
		3958	2nd	11	Kankakee	5692.522	2004	2006	Biological Chemical	54%	46%	36	±15%
		3958	3rd	23	Kankakee	5768.03	2009	2011	Biological Chemical	42%	58%	38	±21%
		3958	4th	33	KankakeeRiver	5768.42	2017	2018	Biological Chemical	41%	59%	38	±14%
Great Lakes Tributaries	04040001 04050001 04100003 04100004 04100005 04100007	4051	1st	5	Great Lakes	5431.33	2000	2002	Biological	20%	80%	38	±10%
		4051	2nd	13	GreatLakes	5431.33	2005	2007	Biological Chemical	42%	58%	37	±14%
		3535	3rd	25	GreatLakes	5688.299	2010	2013	Biological Chemical	18%	82%	37	±11%
		3535	4th	34	GreatLakesTrib	5730.02	2018	2019	Biological Chemical	20%	80%	37	±11%
Ohio River Tributaries	05090203 05140101 05140104 05140201 05140202	3825	1st	5	Ohio River	5273.18	2000	2002	Biological	29%	71%	41	±13%
		3825	2nd	14	OhioRiver	5273.18	2005	2007	Biological Chemical	24%	76%	37	±11%
		3333	3rd	26	OhioRiver	5364.092	2010	2013	Biological Chemical	24%	76%	34	±12%
		3333	4th	35	OhioRiverTrib		2019	2020	Biological Chemical				
Patoka River	05120209	797	2nd	6	Patoka	1080.3	2001	2003	Biological Chemical	37%	63%	26	±15%
		797	3rd	2	PatokaRiver	1150.216	2006	2008	Biological Chemical	28%	72%	37	±14%
		714	4th	27	PatokaRiver	1149.3	2012	2014	Biological Chemical	12%	88%	38	±11%

## Appendix 2 – Summary of locations of Probabilistic Monitoring Program documentation files

Sample Round	Sample Year	My.stratum	My.Type	framesize	Documentation File Name	Documentation File Location	Integrated Report
1st	1996	1	West Fork White River	5974.82			
1st	1997	2	East Fork White River	6802.39			
1st	1997	2	Whitewater	2270.02			
1st	1998	3	Upper Wabash	9548.06			
1st	1999	4	Kankakee	5606.33			
1st	1999	4	Lower Wabash	7601.42			
1st	2000	5	Great Lakes	5431.33			
1st	2000	5	Ohio River	5273.18			
2nd	2001	6	Patoka	1080.3			
2nd	2001	7	WestForkWhiteRiver	5974.82			
2nd	2002	8	Whitewater	2270.018			
2nd	2002	9	EastForkWhite	6802.385			
2nd	2003	10	UpperWabash	9548.056			
2nd	2004	11	Kankakee	5692.522			
2nd	2004	12	LowerWabash	7601.418			
2nd	2005	13	GreatLakes	5431.33			
2nd	2005	14	OhioRiver	5273.18			
3rd	2006	1	WestForkWhiteRiver	6074.201	IN_StreamDesignDocumentation_2006wc.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 3rd Cycle Draw	2010
3rd	2006	2	PatokaRiver	1150.216	IN_StreamDesignDocumentation_2006wc.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 3rd Cycle Draw	2010
3rd	2007	3	EastForkWhiteRiver	6959.485	IN_StreamDesignDocumentation_2006wc.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 3rd Cycle Draw	2010
3rd	2007	4	Whitewater	2330.508	IN_StreamDesignDocumentation_2006wc.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 3rd Cycle Draw	2010
3rd	2008	22	UpperWabash	9904.22	IN_StreamDesignDocumentation_2006wc.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 3rd Cycle Draw	2012
3rd	2009	23	Kankakee	5768.03	IN Stream Design 2006-10 Documentation_update07.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 3rd Cycle Draw	2012
3rd	2009	24	LowerWabash	7474.01	IN_StreamDesignDocumentation_2006wc.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 3rd Cycle Draw	2012
3rd	2010	25	GreatLakes	5688.299	IN Stream Design 2006-10 Documentation_update07.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 3rd Cycle Draw	2014
3rd	2010	26	OhioRiver	5364.092	IN_StreamDesignDocumentation_2006wc.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 3rd Cycle Draw	2014

**Appendix 2 – Summary of locations of Probabilistic Monitoring Program documentation files (cont.)**

Sample Round	Sample Year	My.stratum	My.Type	framesize	Documentation File Name	Documentation File Location	Integrated Report
4th	2011	27	WestForkWhiteRiver	6074.201	IN Stream WestWhite Patoka Design 2011-12 Documentation.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 4th Cycle Draw\WestWhite&Patoka2	2014
4th	2012	27	PatokaRiver	1149.3	IN Stream WestWhite Patoka Design 2011-12 Documentation.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 4th Cycle Draw\WestWhite&Patoka2	2016
4th	2013	28	EastForkWhite	6960.41	IN Stream White Miami 2013-14 Documentation.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 4th Cycle Draw\East Fork White & Great Miami	2016
4th	2014	30	GreatMiamiRiver	2330.09	IN Stream White Miami 2013-14 Documentation.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 4th Cycle Draw\East Fork White & Great Miami	2018
4th	2015	31	UpperWabashRiver	9904.92	IN Stream Wabash 2015-16 Documentation.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 4th Cycle Draw\Upper & Lower Wabash	2018
4th	2016	32	LowerWabashRiver	7625.33	IN Stream Wabash 2015-16 Documentation.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 4th Cycle Draw\Upper & Lower Wabash	2018
4th	2017	33	KankakeeRiver	5768.42	IN Stream 2017-19 Documentation.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 4th Cycle Draw\Kankakee, Great Lakes and Ohio River Tribs	2020
4th	2018	34	GreatLakesTribes	5730.02	IN Stream 2017-19 Documentation.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 4th Cycle Draw\Kankakee, Great Lakes and Ohio River Tribs	2020
4th	2019	35	OhioRiverTribes	5965.23	IN Stream 2017-19 Documentation.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 4th Cycle Draw\Kankakee, Great Lakes and Ohio River Tribs	2022
5th	2020	36	WestForkWhiteRiver	6075.20	IN Stream 2020-22 Documentation.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 5th Cycle Draw\WFWR (2020), Patoka (2021), EFWR (2022)	2022
5th	2021	37	PatokaRiver	1149.30	IN Stream 2020-22 Documentation.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 5th Cycle Draw\WFWR (2020), Patoka (2021), EFWR (2022)	2024
5th	2022	38	EastForkWhiteRiver	6960.41	IN Stream 2020-22 Documentation.pdf	S:\IGCN\OWQ\WSP\OWM\RANDOM\Corvallis 5th Cycle Draw\WFWR (2020), Patoka (2021), EFWR (2022)	2024
5th	2023	39	GreatMiamiRiver				2026
5th	2024	40	UpperWabashRiver				2026
5th	2025	41	LowerWabashRiver				2028
5th	2026	42	KankakeeRiver				2028
5th	2027	43	GreatLakesTribes				2030
5th	2028	44	OhioRiverTribes				2030

### **Appendix 3 – Installing R 3.6.1 and loading the *spsurvey* package.**

R (<https://www.r-project.org/>) is a free, open-source, statistical computing and graphical software environment supported by the R Foundation for Statistical Computing. Although installation of the base R program provides a wide variety of statistical operations, the capabilities of R are greatly increased through the development of more than 15000 user-developed packages providing advanced statistical procedures, graphical user interfaces, or increased graphics production capabilities. Some packages are designed with a very narrow purpose, such as *spsurvey*, (<https://cran.r-project.org/web/packages/spsurvey/index.html>), which provides functions for selecting sites for spatial surveys and analyzing the results of probability survey designs. R is updated periodically with new features; the current version at the time of the writing of this SOP is R 3.6.1, released on 7/5/2019.

R 3.6.1 can be downloaded from a Comprehensive R Archive Network (CRAN) mirror site (<https://cran.r-project.org/mirrors.html>). Click on the link for the nearest CRAN mirror site (i.e. Indiana University, <https://ftp.ussg.iu.edu/CRAN/>), which will take you to a webpage where versions of R are available for download for Linux, Mac, and Windows operating systems. Select Download R for Windows which provides access to four subdirectories from which to choose. Select base which provides you with a link to Download R 3.6.1 for Windows. Run the file R-3.6.1-win.exe to install both the R 3.6.1 32 and 64 bit programs onto your computer and, if given the option, create desktop icons for both versions.

After R 3.6.1 has been installed, open the program. At the top of the R Console window, click on Packages, then Install Package(s). A window opens with a list of CRAN mirrors. Select the nearest mirror site, in this case, USA (MI 1) [https], and click OK. A new window opens with a list of available packages. Scroll down and click on *spsurvey*. Then click OK. The R Console window shows the status of the *spsurvey* package download. After the *spsurvey* package download has completed, at the top of the R Console window click on Packages, then Load Package.... Scroll down and click on *spsurvey*, then click OK. *spsurvey* can also be loaded by typing `library(spsurvey)` into the command line of the R Console and pressing Enter. Either method makes the *spsurvey* package available for use in the current session. Every time R 3.6.1 is restarted, reloading *spsurvey* is required.

## Appendix 4 – R output from the calculation of aquatic life use support estimates for the 2014 Great Miami River Basin Probabilistic Monitoring project

R version 3.6.1 (2019-07-05) -- "Action of the Toes"

Copyright (C) 2019 The R Foundation for Statistical Computing

Platform: i386-w64-mingw32/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.

You are welcome to redistribute it under certain conditions.

Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.

Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

```
> local({pkg <- select.list(sort(.packages(all.available = TRUE)),graphics=TRUE)  
+ if(nchar(pkg)) library(pkg, character.only=TRUE)})
```

Loading required package: sf

Linking to GEOS 3.6.1, GDAL 2.2.3, PROJ 4.9.3

Loading required package: sp

Version 4.1.0 of the spsurvey package was loaded successfully.

```
> framesize <- 2330.09
```

```
> names(framesize) <- 'GreatMiamiRiver'
```

```
> setwd("S:/IGCN/OWQ/WSP/OWM/Integrated Report Data/East Fork White and Whitewater  
Basins/Monitoring Year 2014")
```

```
> DesignStatus <- read.table("designstatus.txt",header = TRUE, sep = "\t", na.strings = "NA", dec = ".",  
strip.white = TRUE)
```

```
> names(DesignStatus)
```

```
[1] "MySite.ID"      "Lsite"          "SiteName"       "Longdd"         "Latdd"          "My.stratum"
```

```

[7] "MDCaty"      "Weight"      "StrahlerOrder"  "My.Type"      "My.Status"
"Overall.Chemistry"

[13] "Fish.Community"  "Fish.IBI"      "Macro.Community"  "Macro.IBI"      "Habitat"
"QHEI"

[19] "DO"          "pH"          "Chloride"        "Nitrate.and.Nitrite"  "Sulfate"

[24] "Lead"        "Copper"      "Zinc"           "Ammonia"        "Cadmium"        "Chromium"

[30] "Nickel"      "Cyanide"     "Aluminum"      "Selenium"       "Nutrients"
"Overall.Chemistry"

[36] "Combined.Assessment"

> tst <- !DesignStatus$My.Status=='NN'

> DesignStatus <- data.frame(DesignStatus[tst,])

> tmp <- marinus(DesignStatus$Latdd,DesignStatus$Longdd)

> DesignStatus$xmarinus <- tmp[, 'x']

> DesignStatus$ymarinus <- tmp[, 'y']

> sites <- rep(TRUE,nrow(DesignStatus))

> DesignStatus$final.wt <- adjwgt(sites,DesignStatus$Weight,DesignStatus$My.Type,framesize)

> sum(DesignStatus$Weight)

[1] 1654.123

> sum(DesignStatus$final.wt)

[1] 2330.09

> write.table(DesignStatus, file="designstatus2.csv",sep=" ",row.names=FALSE)

> designstatus <- read.table('designstatus2.csv',sep=',',header=TRUE)

> sites <- data.frame(siteID=designstatus$MySite.ID,Use=rep(TRUE,nrow( designstatus)))

> subpop <- data.frame( siteID=designstatus$MySite.ID, GreatMiamiRiver=designstatus$My.Type)

> dsgn <- data.frame(
siteID=designstatus$MySite.ID,stratum=designstatus$My.stratum,wgt=designstatus$final.wt,xcoord=de
signstatus$xmarinus,ycoord=designstatus$ymarinus)

> designstatus$StatusTNT <- designstatus$My.Status

> levels(designstatus$StatusTNT) <- list(T=c('TS','SBNC','PB','LD','SCNB','SFNCM','SCMNF','SCFNM'
),NT=c('NT'))

```

```
> data <-  
data.frame(siteID=designstatus$MySite.ID,CombinedStatus=designstatus$My.Status,StatusTNT=designs  
tatus$StatusTNT)  
  
> framelen <- c('GreatMiamiRiver'=2330.09)  
  
> sitestatusum <- cat.analysis(sites, subpop, dsgn, data, popsize=list(GreatMiamiRiver=framelen))  
  
> write.table(sitestatusum, file="sitestatusum2.csv", sep="," ,col.names=NA)  
  
> sites <- data.frame(siteID=designstatus$MySite.ID,Use=designstatus$My.Status=='TS')  
  
> data <- data.frame(siteID=designstatus$MySite.ID,Status=designstatus$Combined.Assessment)  
  
> conditionsum <- cat.analysis(sites,subpop,dsgn,data)  
  
> write.table(conditionsum, file="allconditionsum.csv", sep = " ,",col.names=NA)  
  
> conditionsum90 <- cat.analysis(sites,subpop,dsgn,data,conf=90)  
  
> write.table(conditionsum90, file="allconditionsum90.csv", sep = " ,",col.names=NA)  
  
> savehistory("Great Miami 2014 - R Analyses.txt")  
  
> savehistory("Great Miami 2014 - R Analyses.Rdata")
```